PREDICTING OPERATOR EXECUTION TIMES USING COGTOOL

Yamira Santiago-Espada and Kara A. Latorella, PhD.
NASA Langley Research Center
Hampton, Virginia, 23681

Researchers and developers of NextGen systems can use predictive human performance modeling tools as an initial approach to obtain skilled user performance times analytically, before system testing with users. This paper describes the CogTool models for a two pilot crew executing two different types of a datalink clearance acceptance tasks, and on two different simulation platforms. The CogTool time estimates for accepting and executing Required Time of Arrival and Interval Management clearances were compared to empirical data observed in video tapes and registered in simulation files. Results indicate no statistically significant difference between empirical data and the CogTool predictions. A population comparison test found no significant differences between the CogTool estimates and the empirical execution times for any of the four test conditions. We discuss modeling caveats and considerations for applying CogTool to crew performance modeling in advanced cockpit environments.

The development of Next Generation Aviation flightdeck systems (NextGen) can benefit from the predictive capability of tools that automate the evaluation of operator/interface interaction. Human Performance Models (HPMs) have been shown to play a role in all phases of the concept development, refinement, and deployment process of next generation systems (Gore et al., 2011). A benefit of using HPM tools is that it allows evaluation of conceptual prototypes before a system is built and tested by users.

For over a decade, NASA has been conducting modeling, simulation and human-in-the loop studies to evaluate the impact of diverse NextGen concepts, operations and technologies on pilot performance. One concept investigated since 1970 is the Interval Management with Spacing (IM-S) (Baxley, B., Hubbs, C., Shay, R., Karanian, J., 2011). Under IM-S, the controllers are given support tools that enable them to instruct properly equipped aircraft to manage their own speed to achieve a precise spacing relative to other traffic (Barmore, 2006). A communication protocol based on DataComm will allow the information interchange between ground and aircraft systems.

Since the data available to the pilots on the flightdeck is expected to substantially increase in order to support more precise and closely coordinated operations under NextGen, (Gore, Hooey, Socash, Haan, Mahlstedt, Bawoski…Foyle, 2011, p. 1) it is critical to analyze and understand the ways in which pilots interact with the system. (Medina-Mora, Hoppenbrouwers & Boehm-Davis, 2010). One relevant aspect that is being studied is the data entry tasks performed on the Multi-function control display unit (MCDU). The MCDU is the interface between the pilot and the Flight Management System. Pilots use the MCDU to perform actions such as "monitoring and revising flight plans, selecting operating modes, entering weights, winds, temperatures, as well as initializing performance data" (Honeywell, 2001, p. 2-19).

One such tool, CogTool, has been used to predict execution times of operational procedures for NextGen that involve the use of the MCDU, (John, 2009). CogTool is an interface design evaluation tool that generates predictions based on the Keystroke-Level-Model (KLM) operator

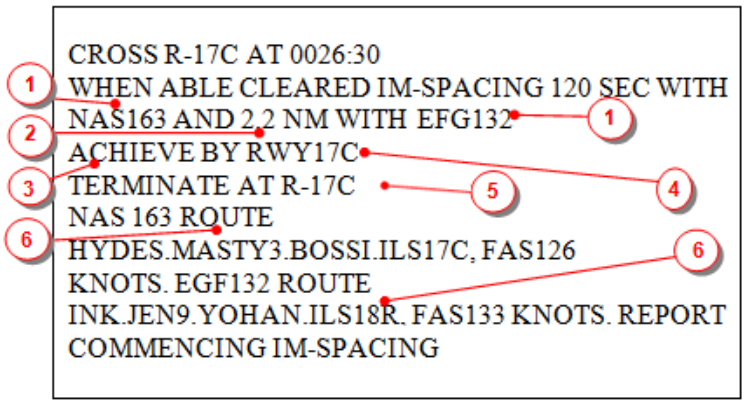
performance model (John, 2011). CogTool estimates of task execution time have been verified empirically in predictive human performance modeling studies (John, 2010).

This paper assesses the effectiveness of CogTool for predicting pilot execution time with a new flight deck interface and crew procedure. CogTool estimates of pilot execution time were compared with empirical data of crews performing Data Comm clearance tasks as part of the Interval Management with Spacing to Parallel Dependent Runways (IMSPiDR1) experiment (Baxley, Hubbs, Shay & Karanian, 2011).

Use of Data Comm messages during the IMSPDiR1 experiment

The IMSPiDR1 experiment included 64 aircraft performing optimized profile descents to the Dallas-Fort Worth terminal area (KDFW). Six out of 64 aircraft were operated by subject pilots, the remaining aircraft were simulated. Of the six aircraft piloted by subject pilots, four of them were flown by single pilots in the Airspace and Traffic Operations simulations stations; one aircraft was piloted by a 2-person crew in the Integration Flight Deck (IFD) and the other crew flew the Development Test (DTS) simulator. The IMSPiDR1 study used the ATOL, IFD and DTS facilities to explore a range of current and future aircraft equipage levels (Baxley et. al, 2011). Pilots followed an Interval Management with Spacing (IM-S) procedure to conduct parallel dependent runway arrival operations. This flight deck procedure employed an onboard avionics system with a spacing tool to provide speed guidance to the flight crew, allowing them to space behind a target aircraft. The flight crews received IM-S clearances through a Controller-Pilot Datalink Communications (CPDLC) with uplink messages issued through the MCDU of the Flight Management System (FMS).

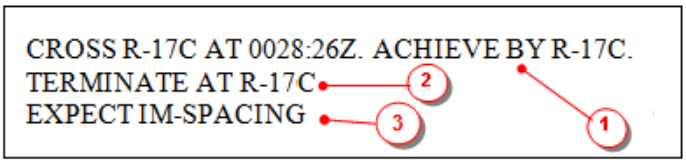
Two types of clearances were issued: a Required Time of Arrival (RTA) clearance and an Interval Management (IM) clearance. Figure 1 shows the information elements of the IM clearance.



Item #	Information element description
1	Target aircraft identification
2	Assigned spacing goal
3	IM clearance type -(achieve then maintain an assigned space interval)
4	Achieve-by-point
5	Termination point
6	Intended flight path information of the target aircraft

Figure 1. An example of a Data Comm "RTA + IM Clearance" message used in IMSPiDR1

The RTA clearance (Figure 2) is brief in comparison to the IM clearance, since it doesn't include the data elements that refer to relative spacing behind the two lead aircraft.



Item #	Information element description
1	Achieve-by-point
2	Termination point
3	Indication to expect IM spacing

Figure 2. An example of a Data Comm "RTA Clearance" message used in IMSPiDR1

Irrespectively of the type of clearance issued by ATC, upon message upload, the crew read the clearance in the MCDU, loaded the information in the Flight-deck Interval Management System, and activated the spacing tool (Swieringa, Murdoch & Baxley, 2011). Once the spacing tool provided the speed to attain the assigned spacing goal, the crew decided if the speed was acceptable. If the speed was acceptable, the crews sent an 'ACCEPT' CPDLC downlink message and executed a route modification in the FMS. The crew monitored the IM-S cues on the PFD and informed ATC when their aircraft was paired with the lead aircraft. Subject pilots flying the IFD, DTS and ATOS stations performed the procedure of clearance reception, verification and acceptance.

Approach to modeling the datalink clearance acceptance in the full-scale simulators

This modeling effort is limited to produce time-on-task estimates that are compared with pilot execution times of task performance in the MCDUs of full mission simulators (IFD and the DTS). The IFD cockpit replicates the flight deck of a Boeing 757-200 aircraft. The DTS is a generic glass cockpit configured with a 777-style Primary Flight Display (PFD) and Navigation Display (ND), a Boeing 767 aisle stand, and a Boeing 757-200 aircraft software model.

The modeling activity began with a review of procedures, manuals and training documents that contained procedural information on how to perform a datalink clearance acceptance task in the IFD and the DTS. The resulting normative model of performance identified three main activities the crew performed in order to accepting a datalink clearance: (1) receiving and loading (L), (2) verifying (V) and (3) accepting (A) a datalink clearance. The modeler also conducted a task analysis of task performance based on video observation. The task analysis characterized these activities in terms of discrete, sequential, and mutually exclusive (non-concurrent) actions performed by 2-person crews - the Pilot Flying and the Pilot Monitoring - to perform the Load-Verify-Accept (LVA) task.

The operational model of execution involved a detailed description of what the crew actually did when performing the LVA task. This process identified actions that pertained to the LVA task, as well as other actions that deviated from the procedures. The modeler reviewed videos and simulation files from IMSPiDR1 to collect human performance measures (operator times) and system response times related to performing the LVA task in the IFD and the DTS. The

calculations of the empirical execution times for the LVA task are based on 51out of 59 data points collected by Baxley et al. On the majority of the 51data points used for analysis, the actions of the crews followed the normative model of task execution. Data points were eliminated where the crews performed the steps out of sequence, forgot to perform steps, or appeared to still be learning the interaction with the MCDU.

The modeler used CogTool's graphical specification capabilities to create storyboards that mock up user interaction on the IFD and DTS interface devices required to accomplish the LVA task. The devices modeled were the Engine Indicating and Crew Alerting System (EICAS) display, and 2 MCDUs. In addition, the modeler included "widgets" for audio input and output that are available in CogTool. Actions on these devices allow the representation of button presses, speaking through the microphone, hearing and reading.

The majority of the steps pertaining to the LVA task were similar; however, there were subtle differences in the representation of the 'LOAD' and 'ACCEPT' task, due to the variation in the length of the clearance. The representation of the Pilot Monitoring reading the clearance involved more 'Look-at' operations on the IM clearance than in the RTA clearance. In general, the acceptance of an IM clearance involved more 'NEXT PAGE' button presses than the RTA clearance. The modeler added a delay of 17.7s in the CogTool model of the Verify subtask to account for the time that it took the onboard tool to provide speed guidance. To simulate the actions of 2-person crew, a one second delay was inserted before the Pilot Flying executed the command to represent the time awaiting concurrence from the Pilot Monitoring.

After specifying the IFD and DTS, the modeler used the mouse to represent the execution of the LVA task on the mockups of the IFD and DTS flight decks. CogTool recorded the actions, produced a task analysis and a prediction of the time required to complete the task, based on a human performance model. The CogTool time estimates were compared with empirical times obtained from a sample of IMSPiDR data. The modeler performed statistical significance tests comparing the empirical median execution times with the CogTool estimate.

Results

C**ogTool prediction on time to perform the LVA task for datalink clearance acceptance**

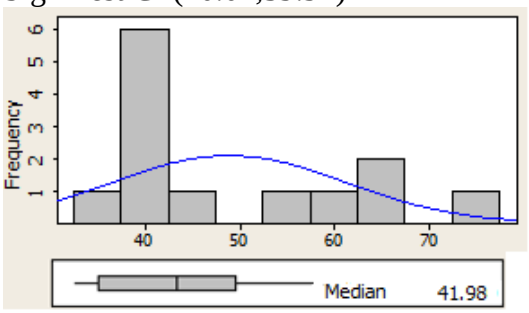
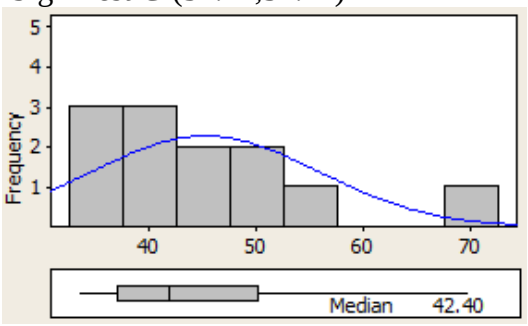
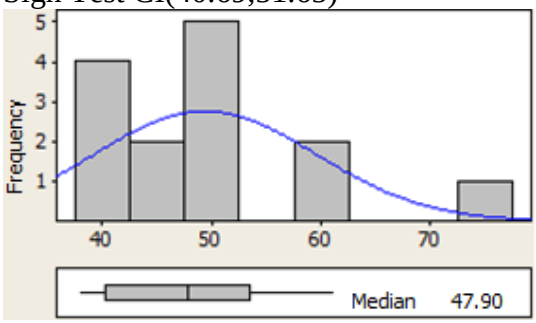
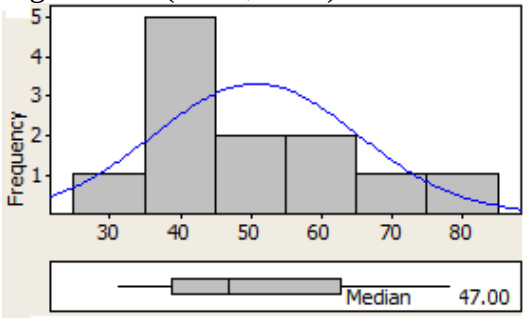
CogTool predicted a total execution time of 46.96s to perform the RTA clearance acceptance task in the IFD and 46.33s on the DTS. For the IM-s clearance acceptance task, CogTool predicted a total execution time of 49.52s on the IFD and 48.96s on the DTS.

Empirical Execution times for clearance Acceptance on the IFD and DTS

The median times for the actual datalink clearance acceptance observed on the IFD were 41.98s for the RTA clearance (RTA-IFD), and 47.90s for the IM clearance (IM-IFD). The median times of execution of the actual datalink clearance acceptance observed on the DTS were 42.40s for the RTA clearance (RTA-DTS) and 47.0s for the IM clearance (IM-DTS). Table 2 shows the mean execution times of the datalink clearance acceptance task, and the distributions of the data.

The mean times were greater than the median times for the RTA and IM clearances in the IFD and DTS (Table 2).

Table 2. Distribution of the data, CogTool estimates and Confidence intervals for the Sign Test of median execution times of datalink clearance acceptance.

RTA clearance	IFD CogTool estimate= 45.97 Sign Test CI (40.02,59.92)  n= 13	DTS CogTool estimate = 46.33 Sign Test CI(37.41,51.17)  n=12
IM clearance	IFD CogTool estimate= 49.52 Sign Test CI(40.69,51.65)  n=14	DTS CogTool estimate= 48.96 Sign Test CI(38.99,62.74)  n=12

Comparing the distribution medians with the CogTool estimate

The Signed Rank Test ascertains if there were significant differences between the CogTool estimates and the empirical data for both types of clearances. The significance of the test is given by the p-values (p). Table 2 shows the Confidence Intervals of the Signed Test at the 0.05 level of significance. The CogTool estimate of task execution time for the RTA clearance acceptance on the IFD (45.97s) is not significantly different from the median empirical execution times (p=0.5811). Similarly, there were not significant differences between the predicted execution time for the IM Clearance (49.52s) and the median empirical execution times for subjects flying in the IFD (p=0.4240).

The CogTool estimates of task execution time for the RTA clearance acceptance on the DTS (46.33s), is not significantly different from the median empirical execution time (p=0.3877).

There are not significant differences between the CogTool estimate of task execution time for the IM Clearance (48.96 s) and the empirical execution times (p=1).

The modeler observed conditions which accounted for the spread in empirical data and therefore imprecision of the CogTool estimates with respect to it. In 31 out of the 51 data samples from IMSPiDR1, the crew execution time was less than the CogTool estimate. Among the reasons for which CogTool might have overestimated the time to perform the LVA task are the following:

1. In the CogTool model, the Pilot Monitoring (PM) briefs the clearance to the Pilot Flying (PF) using the exact words that appear in the CPDLC uplink. While subject pilots of IMSPiDR1 verbalized the clearance (except in one case), they also paraphrased or shortened the briefing with phrases such as "Here NASA 094 and Eagle again", or "Same as before".
2. Actions that were performed in parallel in IMS-PiDR1 were represented sequentially in CogTool. In CogTool the actions of reading the clearance and briefing the clearance can't be done in parallel. During the experiment, the PM briefs the PF while reading the clearance.
3. The model assumed independence (mutual exclusiveness) of actions between crew members, when there was actually collaboration. It was observed that in a few instances one crew member made the steps intended for the other member; or used two MCDU's.
4. There were at least 3 instances where the work distribution was different from what was prescribed in the procedures. In a few cases, the PM used both MCDU's to look the clearance information; while he read the information in one MCDU, he accepted the clearance using the other MCDU.

In 20 out of 51 data samples, the crew execution time was greater than the CogTool estimate. Among the reasons for which CogTool might have underestimated the time to perform the LVA task are the following:

1. The model assigned mental operators that are associated with observable actions of user interaction, such as looking, listening and pointing objects; however, it doesn't capture mental operations related to complex thinking. The modeler used one 'think' operation to represent the process of deciding whether a clearance is acceptable.
2. In a few instances, crews performed extra steps to verify that the information given on the clearance matched the existing information on the Flight Progress.
3. Crew performed additional, parallel tasks during the clearance reception; e.g. PM that replies back to ATC while loading the clearance.
4. The CogTool model of the LVA task didn't take into account the delay caused by pilots focusing their visual attention on a region of the MCDU before locating a specific button. John et al. (2009) states that a model of using a complex device should choose a region of the device to focus attention on rather than treating all the buttons as equally possible at all times.

5. The CogTool model assumed expert performance; however, Baxley et al. reported that the execution times greater than 60 seconds were associated with equipment or procedural questions.

Conclusions

CogTool estimates are representative of the actual execution times of pilots performing the datalink clearance acceptance tasks in IMSPiDR1. The CogTool predictions didn't significantly differ from the times obtained from the random samples obtained from the IMSPiDR1 experiment. CogTool is suitable to obtain estimates on expert performance on frequent tasks that can be described in terms of discrete actions, such as user interaction with the MCDU. However, oversimplifications on modeling pilot interaction, either with elements of the flight deck or another crew member, introduce inaccuracies in the time estimates of the predictive tool.

The CogTool model of the datalink clearance acceptance can be tailored by adding user delays. Those delays account for the times in which the crews are performing other actions that were not represented, but that are part of deciding whether or not the clearance is acceptable. One of such actions is looking at the speed guidance information in the Primary Flight Display. In order to improve the CogTool's time estimates is necessary to make adjustments to the model to do a better representation of the reading of the clearance. The modeling of reading the text of the clearances in CogTool utilized a sequence of 'Look-at' operators. CogTool assigns an execution time of 1.2 seconds to the Look-at operator. A tailored model to represent the time that takes a pilot to read a clearance would use only one 'Look-at' operator with a time value of 5.9 seconds, instead of the default 1.2 seconds. This approximation of 5.9 seconds to read a clearance is based on the results from the IMSPiDR study, as reported by Baxley et al. The CogTool representation of modeling reading the clearance in the MCDU can also be attained by prototyping a two-level visual search by focusing first on a region of the MCDU and then on the specific buttons in that region(John, Blackmon, Polson, Fennel, Teo, 2009).

CogTool's functionalities have been extended or combined with other automated usability design tools in order to modeling exploratory behavior in complex devices. Teo and John (2008) implemented a model for visual search in a version of CogTool suitable to generate metrics of non expert performance. CogTool Explorer, an add-on to CogTool, predicts exploratory behavior exhibited by novice users when presented with a new interface (Teo & John, 2008). Sherry, John & Polson (2010) reported on research effort that combined CogTool with the Human Interaction Process Analysis model (HCIPA) operator performance models and the Aviation Cognitive Walkthrough task specification language to analyze programming tasks in the MCDU of a Boeing 777.

Automation design tools that incorporate human performance models are useful to analyze skilled performance on routine interactive tasks performed in the cockpit. When tailored to model the behavior under study on a prototype of the system, automation design tools can help designers to understand the ways in which pilots interact with the cockpit automation, before bringing the subjects to test the system. The employment of automation design tools in an early stage of the life-cycle of a system can be an effective way of reducing its design cost (Sherry, John, Polson, 2010). As the robustness of these tools increase, so does their inclusion in usability analysis of

systems design; and with that, the expectation that their utilization would have a positive impact on the research, design and development of new systems.

Acknowledgements

The authors would like to thank Bryan Baxley, Bryan Barmore, Jennifer Murdoch and Kurt Swieringa for providing the documentation and results from IMSPiDR1 experiment useful for the development of the model and analysis of the data. The authors appreciate the help of Paul Sugden and Miguel Álvarez for supplying information regarding the equipment and software used in the IFD and DTS facilities.

References

Barmore, B. (2006). Airborne precision spacing: A trajectory-based approach to improve terminal area operations. *25th Digital Avionics Systems Conference, Portland, Oregon, October 15-19, 2006.* http://dx.doi.org/10.1109/DASC.2006.313683

Baxley, B. T., Hubbs, C., Shay, R., & Karanian, J. (2011). Use of data comm by flight crew to conduct interval management operations to parallel dependent runways. In *11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Virginia Beach, VA, March 20-22 2013: American Institute of Aeronautics and Astronautics (AIAA).* http://dx.doi.org/10.2514/MATIO11

Gore, B. F, Hooey, B.L, Socash, C.M., Haan, N.J., Mahlstedt, E, Bakowski, D.L., …Foyle, D.C. (2011). *Evaluating NextGen closely space parallel operations concepts with validated human performance models. NASA Ames Research Center, Human System Integration Division*. CA. Retrieved from http://humansystems.arc.nasa.gov/groups/midas/documents/HCSL_Tech_Report_HCSL_11_01.pdf

Honeywell. (2011). *Boeing 777 Flight Management System Pilot's Guide.* (Publication No. C28--3641--022--01).

John, B. E. (2010) Reducing the variability between novice modelers: Results of a tool for human performance modeling produced through Human-Centered Design. *Proceedings of the 19th Annual Conference on Behavior Representation in Modeling and Simulation (BRIMS), Charleston, South Carolina, March 22-25, 2010.*

John, B. E. (2011). A quantitative usability assessment method for inclusion in software engineering courses. *Proceedings of the 2011 24th IEEE-CS Conference on Software Engineering Education and Training* (pp.537-539). http://dx.doi.org/10.1109/CSEET.2011.5876143

John, B. E., Blackmon, M. H., Polson, P. G., Fennell, K., & Teo, L. H. (2009). Rapid theory prototyping: An example of an aviation task. *Proceedings of the Human Factors and Ergonomics Society the 53rd Annual Meeting, Santa Monica, CA, October 19-23, 2009*

Swieringa, K., Murdoch, J. L. Baxley, B. (2011). Evaluation of an airborne spacing concept, on-board spacing tool, and pilot interface. *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference: Virginia Beach, VA*, September *20-22, 2011,* American Institute of Aeronautics and Astronautics (AIAA).

Teo, L. & John, B. E. (2008) Towards Predicting User Interaction with CogTool-Explorer. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting (New York City, New York, Sept 22-26, 2008*.